



**HAL**  
open science

# Heaviness in a Verb-final Language: Evidence from Persian

Pegah Faghiri, Pollet Samvelian, Barbara Hemforth

► **To cite this version:**

Pegah Faghiri, Pollet Samvelian, Barbara Hemforth. Heaviness in a Verb-final Language: Evidence from Persian. AMLaP 2013, Sep 2013, Marseille, France. 2013. halshs-01131210

**HAL Id: halshs-01131210**

**<https://shs.hal.science/halshs-01131210>**

Submitted on 13 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Heaviness in a Verb-final Language: Evidence from Persian

Pegah Faghiri<sup>1,2</sup>, Pollet Samvelian<sup>1</sup> & Barbara Hemforth<sup>2</sup>

<sup>1</sup> Université Paris 3 - Mondes iranien et indien (CNRS), <sup>2</sup> Université Paris 7 - Laboratoire de Linguistique Formelle (CNRS)

pegah.faghiri@univ-paris3.fr



## Introduction

- Is the **short-before-long principle** universal?
  - Heavy constituents require more processing resources (e.g. Arnold *et al*, 2000 ; Stallings *et al*, 1998)
  - Costly constituents tend to be postponed
- Meanwhile, Hawkins's *Early Immediate Constituent (EIC)* principles predicts opposite tendencies
  - For head-initial and head-final languages (Hawkins, 1990 a.o.)
  - **Long-before-short** in head-final languages
  - Confirmed for Japanese by corpus and experimental data (Hawkins, 1994 ; Yamashita & Chang, 2001)

**EIC principle in Japanese**

[Mary-ga]	[kinoo	John-ga	kekconsi-ta	to]	it-ta	
Mary-NOM	yesterday	John-NOM	married	that	said	
Mary said that John got married yesterday.						
[Mary-ga]	[kinoo	John-ga	kekconsi-ta	to]	it-ta	
1	2	3	4	5	6	
[kinoo	John-ga	kekconsi-ta	to]	[Mary-ga]	it-ta	
1	2	3	4	5	6	

DO IO or IO DO ?  
PP > NP (by 2 words)

[NP]	[NP]	[P]	V
1	2-3	4	5
[NP]	[P]	[NP]	V
1	2	3	4

## Persian: Essential Properties

- Mixed head direction:
    - Head-final in verbal domain (SOV)
    - Head-initial elsewhere
    - e.g. Det N Mod, Prep NP, Comp P
  - EIC does not (always) work for Persian:
    - NP PP V
    - PP NP V

No preferential order is predicted by the EIC principle
  - Canonical word order is SOV
  - But variation is possible (SVO, VSO, etc.)
    - Depending on register, information structure, prosody, etc.
    - e.g. Goal arguments are post-verbal in oral/informal register
  - Clausal complements are strictly post-verbal
  - Differential Object Marking (DOM)
    - A definite and/or specific DO is always marked with =*rā* (cf. ex 4)
    - An indefinite non-specific DO is **not** marked (cf. ex 1 – 3)
  - Most prominent hypothesis for complement ordering is the DOM criterion
    - Marked DOs can be separated from the verb → NP PP V order
    - Unmarked DOs should be adjacent to the verb → PP NP V order (Karimi, 2005 a.o.)
  - Part of this hypothesis has been shown not to hold in corpus data (Faghiri & Samvelian, 2013)
- Object of this study:**  
**Preferential order between the DO and the IO in the preverbal domain**

## Corpus Study

Corpus:  
Bijankhan corpus : 2,6m tokens, from newspapers, annotated for POS, freely available  
Verbs were lemmatized and potentially ditransitive verbs were extracted (42k tokens, 122 types)

Dataset (908 tokens, 82 lemmas):  
Selection of sentences out of

1. A random sample of 2000 tokens: 541 occ.
2. All instances of verbs 'to send' and 'to pour' (low frequency)
3. Random samples of verbs 'to give' and 'to take' (very high frequency)

Mixed-effect logistic regression  
Dependent variable: Order (NP PP V = 1)  
Fixed effects:  
1. DO type  
2. Relative length (nb of words): log(NP) – log(PP)  
Random effect: Verb lemma

**Average preference of 59% for NP-PP-V order**

4 DO types (based on preliminary observations)

- Bare**
- (1) Maryam *be* Nima *ketāb* dād  
Maryam to Nima book gave  
'Maryam gave a book/books to Nima.'
- Bare-Modified**
- (2) Maryam *be* Nima *ketāb=e tārix* dād  
Maryam to Nima book=EZ\* history gave  
'Maryam gave a history book/history books to Nima.'
- Indefinite**
- (3) Maryam *čand ketāb=e qadimi* *be* Nima dād  
Maryam some book=EZ old to Nima gave  
'Maryam gave some old books to Nima.'
- Marked (= *rā*)**
- (4) Maryam *in ketāb=rā* *be* Nima dād  
Maryam this book=DOM to Nima gave  
'Maryam gave this book to Nima.'
- DO type determines the relative order in 87% of cases

	Bare	Bare-Modified	Indefinite	Marked
NP PP V	43 (16%)	23 (34%)	112 (77%)	404 (95%)
PP NP V	228 (84%)	44 (66%)	33 (23%)	21 (5%)
Total	271	67	145	425

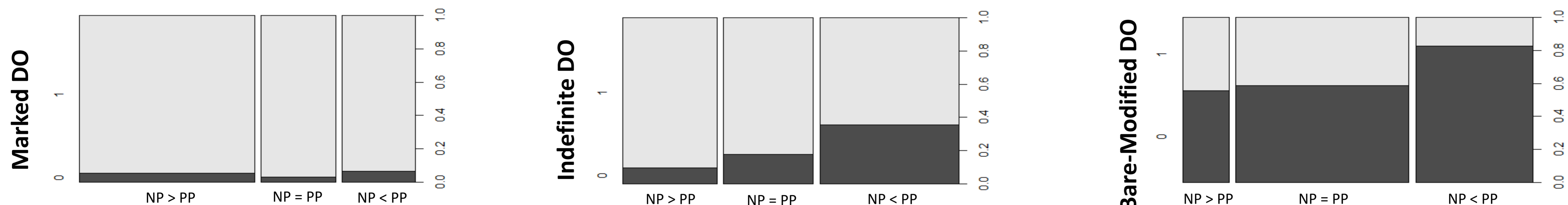
Relative length is irrelevant for Bare DOs  
NP ≤ PP in all cases

The relative-length plays a role in the choice of the relative order in the case of these DO types

## The Relative Length

Beyond the strong effect of DO type, **relative length** shows a significant effect (p-value < 0.001) corresponding to the **long-before-short** tendency

Shorter NPs prefer the PP NP V order significantly more often



→ Persian behaves like Japanese, confirming the long-before-short principle in OV languages

## Discussion and Conclusions

- The short-before-long principle is not universal:
  - Not only Japanese (strictly head-final) but also Persian (mixed head-direction) presents the long-before-short tendency
  - The verbal position has to be taken into account in the effect of relative length on preferential order between verbal complements
  - Theories solely based on general principles ignoring linguistic parameters would eventually fail cross-linguistic validity
  - Theories proposing accounts in terms of dependency seems to be more appropriate
  - ❖ However Hawkins's *EIC* principles fails to account for Persian data
- In Persian the relative length plays only a secondary role while the DO type, which depends on the information status of the NP, plays the essential role
- We are currently running a number of experiments to explore the effect of information structure and the relative length independently